ELSEVIER

Journal of Hydrology 228 (2000) 113–129

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall

P. Goovaerts*

Department of Civil and Environmental Engineering, The University of Michigan, Ann Arbor, MI 48109-2125, USA

Received 11 February 1999; received in revised form 11 November 1999; accepted 16 December 1999

Abstract

This paper presents three multivariate geostatistical algorithms for incorporating a digital elevation model into the spatial prediction of rainfall: simple kriging with varying local means; kriging with an external drift; and colocated cokriging. The techniques are illustrated using annual and monthly rainfall observations measured at 36 climatic stations in a 5000 km^2 region of Portugal. Cross validation is used to compare the prediction performances of the three geostatistical interpolation algorithms with the straightforward linear regression of rainfall against elevation and three univariate techniques: the Thiessen polygon; inverse square distance; and ordinary kriging.

Larger prediction errors are obtained for the two algorithms (inverse square distance, Thiessen polygon) that ignore both the elevation and rainfall records at surrounding stations. The three multivariate geostatistical algorithms outperform other interpolators, in particular the linear regression, which stresses the importance of accounting for spatially dependent rainfall observations in addition to the colocated elevation. Last, ordinary kriging yields more accurate predictions than linear regression when the correlation between rainfall and elevation is moderate (less than 0.75 in the case study).

Keywords: Rainfall; DEM; Multivariate geostatistics; Kriging

1. Introduction

Measured rainfall data are important to many problems in hydrologic analysis and designs. For example the ability of obtaining high resolution estimates of spatial variability in rainfall fields becomes important for identification of locally intense storms which could lead to floods and especially to flash floods. The accurate estimation of the spatial distribution of rainfall requires a very dense network of instruments, which entails large installation and operational costs. Also, vandalism or the failure of the observer to make the necessary visit to the gage may result in even lower sampling density. It is thus necessary to estimate point rainfall at unrecorded locations from values at surrounding sites.

A number of methods have been proposed for the interpolation of rainfall data. The simplest approach consists of assigning to the unsampled location the record of the closest gage (Thiessen, 1911). This method amounts at drawing around each gage a polygon of influence with the boundaries at a distance halfway between gage pairs, hence the name Thiessen polygon for the technique. Although the Thiessen polygon method is essentially used for estimation of areal rainfall (McCuen, 1998), it has also been applied to the interpolation of point measurements (Creutin

* Tel.: +1-734-936-0141; fax: +1-734-763-2275.
E-mail address: goovaert@engin.umich.edu (P. Goovaerts).

0022-1694/00/$ - see front matter
PII: S0022-1694(00)00144-X

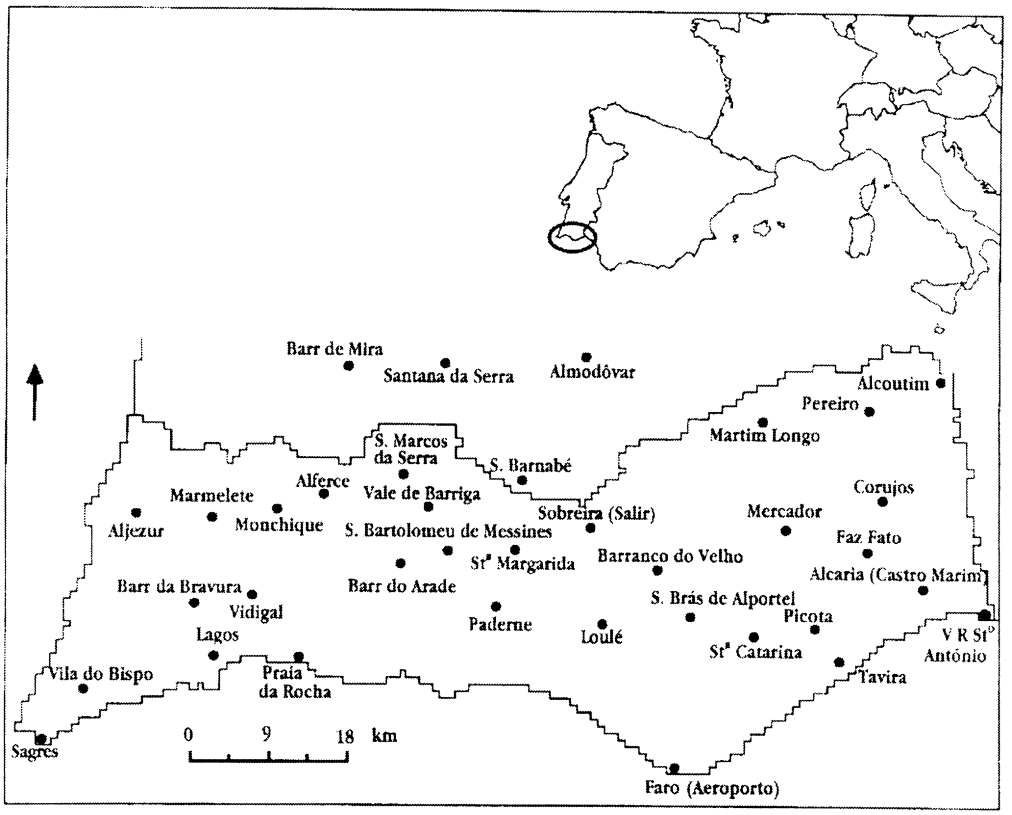

Fig. 1. Location of the study area and positions of the 36 climatic stations.

and Obled, 1982; Tabios and Salas, 1985; Dirks et al., 1998). In 1972, the US National Weather Service has developed another method whereby the unknown rainfall depth is estimated as a weighted average of surrounding values, the weights being reciprocal to the square distances from the unsampled location (Bedient and Huber, 1992, p. 25). Like the Thiessen polygon method, the inverse square distance technique does not allow the hydrologist to consider factors, such as topography, that can affect the catch at a gage. The isohyetal method (McCuen, 1998, p. 190) is designed to overcome this deficiency. The idea is to use the location and catch for each gage, as well as knowledge of the factors affecting these catches, to draw lines of equal rainfall depth (isohyets). The amount of rainfall at the unsampled location is then estimated by interpolation within the isohyets. A limitation of the technique is that an extensive gage network is required to draw isohyets accurately.

Geostatistics, which is based on the theory of regionalized variables (Journel and Huijbregts, 1978; Goovaerts, 1997, 1999), is increasingly preferred because it allows one to capitalize on the spatial correlation between neighboring observations to predict attribute values at unsampled locations. Several authors (Tabios and Salas, 1985; Phillips et al., 1992) have shown that the geostatistical prediction technique (kriging) provides better estimates of rainfall than conventional methods. Recently, Dirks et al. (1998) found that the results depend on the sampling density and that, for high-resolution networks (e.g. 13 raingages over a 35 km^2 area), the kriging method does not show significantly greater predictive skill than simpler techniques, such as the inverse square distance method. Similar results were found by Borga and Vizzaccaro (1997) when they compared kriging and multiquadratic surface fitting for various gage densities. In fact, besides providing a measure of

Table 1
Statistics for the monthly and annual rainfall data (36 observations). The last three columns give the linear correlation coefficient between rainfall and elevation, and the mean absolute error (MAE) and mean square error (MSE) of prediction of rainfall by linear regression of elevation

Period	Rainfall (mm)				Correlation	MAE	MSE
	Mean	Std. dev.	Min.	Max.			
Jan	69.9	24.7	37.6	137.0	0.69	13.9	317
Feb	58.0	25.7	27.4	146.6	0.75	12.5	287
Mar	32.7	11.4	17.2	73.9	0.80	5.05	47.4
Apr	42.1	17.3	17.9	105.9	0.74	8.91	135
May	21.0	9.8	7.2	54.6	0.83	4.30	30.1
Jun	8.1	3.7	3.2	16.8	0.83	1.71	4.33
Jul	1.0	0.8	0.0	3.2	0.39	0.66	0.61
Aug	2.0	1.2	0.0	5.3	0.33	0.85	1.21
Sep	12.1	4.1	5.6	22.8	0.75	2.24	7.56
Oct	59.6	16.1	32.2	111.0	0.76	8.38	111
Nov	79.3	22.0	39.4	148.7	0.72	11.1	231
Dec	96.3	33.7	44.4	183.0	0.71	19.9	562
Annual	482.1	159.8	259.5	1005	0.79	76.2	9774

prediction error (kriging variance), a major advantage of kriging over simpler methods is that the sparsely sampled observations of the primary attribute can be complemented by secondary attributes that are more densely sampled. For rainfall, secondary information can take the form of weather-radar observations. A multivariate extension of kriging, known as cokriging, has been used for merging raingage and radar-rainfall data (Creutin et al., 1988; Azimi-Zonooz et al., 1989). Raspa et al. (1997) used another geostatistical technique, kriging with an external drift, to combine both types of information. In this paper, another valuable and cheaper source of secondary information is considered: digital elevation model (DEM). Precipitation tends to increase with increasing elevation, mainly because of the orographic effect of mountainous terrain, which causes the air to be lifted vertically, and the condensation occurs due to adiabatic cooling. For example Hevesi et al. (1992a,b) reported a significant 0.75 correlation between average annual precipitation and elevation recorded at 62 stations in Nevada and southeastern California. In their paper, they used a multivariate version of kriging, called cokriging, to incorporate elevation into the mapping of rainfall. A more straightforward approach consists of estimating rainfall at a DEM grid cell through a regression of rainfall versus elevation (Daly et al., 1994).

In this paper, annual and monthly rainfall data from the Algarve region (Portugal) are interpolated using two types of techniques: (1) methods that use only rainfall data recorded at 36 stations (the Thiessen polygon, inverse square distance, and ordinary kriging); and (2) algorithms that combine rainfall data with a digital elevation model (linear regression, simple kriging with varying local means, kriging with an external drift, colocated ordinary cokriging). Prediction performances of the different algorithms are compared using cross validation and are related to the strength of the correlation between rainfall and elevation, and the pattern of spatial dependence of rainfall.

2. Case study

The Algarve is the most southern region of Portugal, with an area of approximately 5000 km^2. Fig. 1 shows the location of 36 daily read raingage stations used in this study. The monthly and annual rainfall depths have been averaged over the period of January 1970–March 1995, and basic sample statistics (mean, standard deviation, minimum, maximum) are given in Table 1. The subsequent analysis will be conducted on these averaged data, hence fluctuations of monthly and annual precipitations from one year to another will not be investigated.

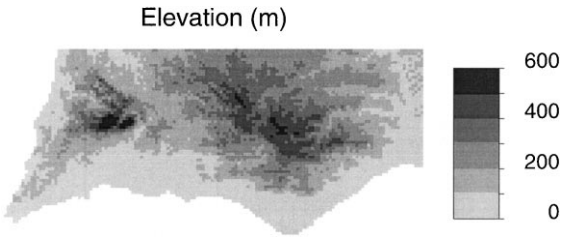

Fig. 2. Digital elevation model.

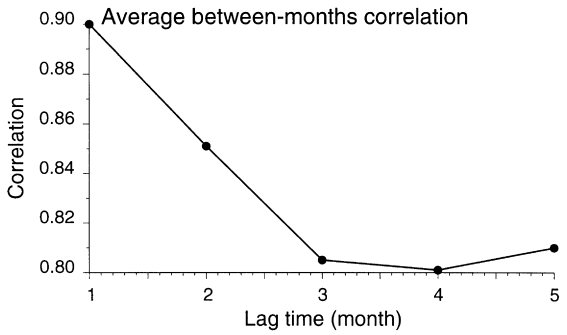

Fig. 3. Average correlation between monthly rainfall data measured at increasing time intervals: 1–5 months.

Another source of information is the elevation map shown in Fig. 2. Each grid cell represents 1 km^2 and its elevation was computed as the average of the elevations at 4 discrete points within the cell. The relief is dominated by the two main Algarve's mountains: the Monchique (left) and the Caldeirão (right). Table 1 indicates that the correlation between rainfall and elevation ranges from 0.33 to 0.83, hence it seems worth accounting for this exhaustive secondary information into the mapping of rainfall. The control of elevation on the spatial distribution of rainfall explains the moderate to strong correlation between monthly rainfall data, see Table 2. Apart from the two dry months of July and August the correlation ranges from 0.50 to 0.97. The correlation coefficients of Table 2 have been averaged as a function of the time lag between months, excluding July and August. Fig. 3 shows that the average correlation between rainfall measured over two consecutive months (lag = 1) is 0.9 and slightly decreases to 0.8 as the separation time increases.

3. Interpolation procedures

This section briefly introduces the different estimators used in the case study. Interested readers should refer to Goovaerts (1997) for a detailed presentation of the different kriging algorithms, and Deutsch and Journel (1998) for their implementation in the public-domain Geostatistical Software Library (Gslib).

3.1. Univariate estimation

Consider first the problem of estimating the rainfall depth z at an unsampled location $\mathbf{u}$ using only rainfall data. Let $\{z(\mathbf{u}_\alpha), \alpha = 1,\ldots,n\}$ be the set of rainfall data measured at $n = 36$ locations $\mathbf{u}_\alpha$.

The most straightforward approach is the Thiessen

Table 2
Matrix of linear correlation coefficients between monthly rainfall data (36 observations)

Period	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	1.00	0.97	0.86	0.85	0.84	0.78	0.23	0.31	0.58	0.88	0.94	0.94
Feb		1.00	0.92	0.91	0.87	0.81	0.22	0.42	0.64	0.92	0.96	0.93
Mar			1.00	0.96	0.89	0.80	0.20	0.47	0.71	0.89	0.89	0.81
Apr				1.00	0.91	0.82	0.20	0.42	0.71	0.91	0.89	0.78
May					1.00	0.88	0.33	0.36	0.75	0.90	0.85	0.78
Jun						1.00	0.38	0.34	0.77	0.84	0.79	0.72
Jul							1.00	−0.04	0.26	0.12	0.24	0.34
Aug								1.00	0.46	0.37	0.35	0.31
Sep									1.00	0.75	0.65	0.50
Oct										1.00	0.89	0.81
Nov											1.00	0.91
Dec												1.00

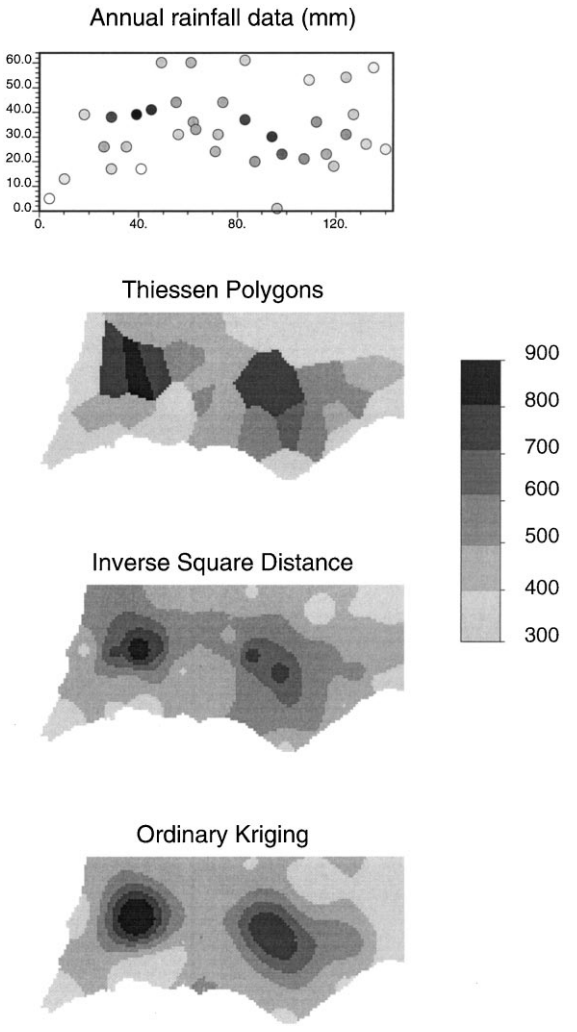

Fig. 4. Annual rainfall maps obtained by interpolation of 36 observations (top map) using the Thiessen polygon, inverse square distance, and ordinary kriging.

polygon method whereby the value of the closest observation is simply assigned to **u**:

$$z^*_{\mathrm{Pol}}(\mathbf{u}) = z(\mathbf{u}_{\alpha'}) \text{ with } |\mathbf{u}-\mathbf{u}_{\alpha'}| < |\mathbf{u}-\mathbf{u}_{\alpha}| \ \forall \alpha \neq \alpha'. \quad (1)$$

Fig. 4 (2nd graph) shows the map of annual rainfall interpolated at the nodes of a $1 \times 1\ \mathrm{km}^2$ grid corresponding to the resolution of the elevation model. The map displays the characteristic polygonal zones of influence around the 36 gages.

To avoid unrealistic patchy maps, the depth z can be estimated as a linear combination of several surrounding observations, with the weights being inversely proportional to the square distance between observations and **u**:

$$z^*_{\mathrm{Inv}}(\mathbf{u}) = \frac{1}{\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})} \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) z(\mathbf{u}_\alpha) \quad (2)$$

$$\text{with } \lambda_\alpha(\mathbf{u}) = \frac{1}{|\mathbf{u}-\mathbf{u}_\alpha|^2}$$

Fig. 4 (3rd graph) shows the map of annual rainfall produced using the inverse square distance method and $n(\mathbf{u}) = 16$ surrounding observations.

The basic idea behind the weighting scheme (2) is that observations that are close to each other on the ground tend to be more alike than those further apart, hence observations closer to **u** should receive a larger weight. Instead of the Euclidian distance, geostatistics uses the semivariogram as a measure of dissimilarity between observations. The experimental semivariogram $\hat{\gamma}(\mathbf{h})$ is computed as half the average squared difference between the components of data pairs:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})]^2, \quad (3)$$

where $N(\mathbf{h})$ is the number of pairs of data locations a vector **h** apart. The semivariogram is a function of both the distance and direction, and so it can account for direction-dependent variability (anisotropic spatial pattern).

Fig. 5 (top graph) shows the semivariogram of annual rainfall computed from the 36 data of Fig. 4. Because of the lack of data only the omnidirectional semivariogram was computed, and hence the spatial variability is assumed to be identical in all directions. Semivariogram values increase with the separation distance, reflecting our intuitive feeling that two rainfall data close to each other on the ground are more alike, and thus their squared difference is smaller, than those that are further apart. The semivariogram reaches a maximum at 25 km before dipping and fluctuating around a sill value. The so-called "hole effect" typically reflects pseudo-periodic or cyclic phenomena (Journel and Huijbregts, 1978, p. 403). Here, the

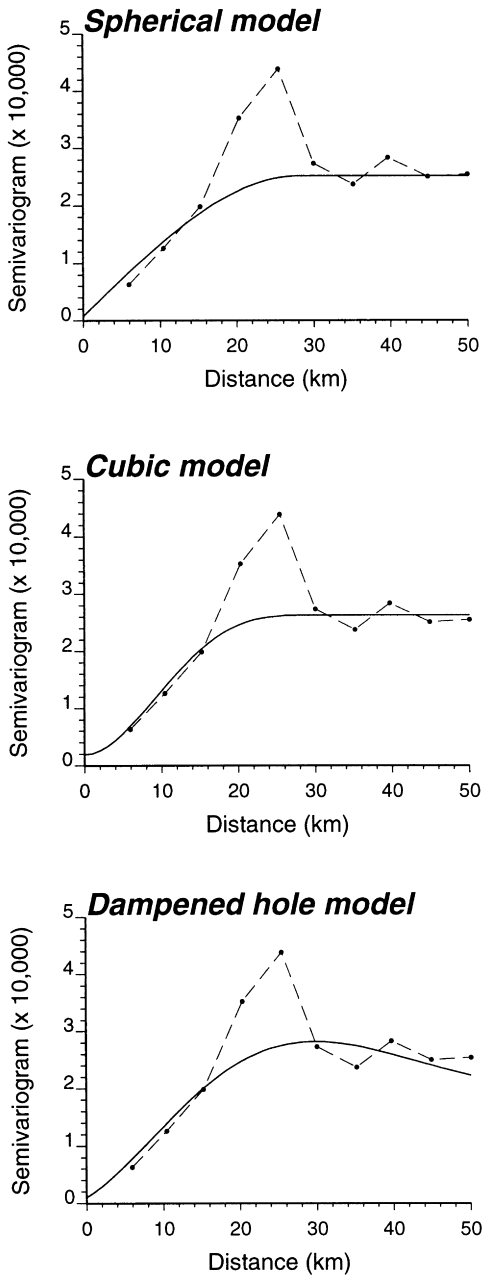

Fig. 5. Experimental semivariogram of the annual rainfall with three different permissible models fitted.

hole effect relates to the existence of two mountains 40 km apart (recall Fig. 2) which creates two high-valued areas in the rainfall field.

Kriging is a generalized least-square regression technique that allows one to account for the spatial dependence between observations, as revealed by the semivariogram, into spatial prediction. Most of geostatistics is based on the concept of a random function, whereby the set of unknown values is regarded as a set of spatially dependent random variables. Each measurement $z(\mathbf{u}_\alpha)$ is thus interpreted as a particular realization of a random variable $Z(\mathbf{u}_\alpha)$. Interested readers should refer to textbooks such as Isaaks and Srivastava (1989, pp. 196–236) or Goovaerts (1997, pp. 59–74) for a detailed presentation of the theory of random functions. Like the inverse square distance method, geostatistical interpolation amounts at estimating the unknown rainfall depth z at the unsampled location $\mathbf{u}$ as a linear combination of neighboring observations:

$$z^*_{\mathrm{OK}}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{\mathrm{OK}}_\alpha(\mathbf{u}) z(\mathbf{u}_\alpha) \quad \text{with} \quad \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{\mathrm{OK}}_\alpha(\mathbf{u}) = 1. \quad (4)$$

The ordinary kriging weights $\lambda^{\mathrm{OK}}_\alpha(\mathbf{u})$ are determined such as to minimize the estimation variance, $\mathrm{Var}\{Z^*_{\mathrm{OK}}(\mathbf{u}) - Z(\mathbf{u})\}$, while ensuring the unbiasedness of the estimator, $\mathrm{E}\{Z^*_{\mathrm{OK}}(\mathbf{u}) - Z(\mathbf{u})\} = 0$. These weights are obtained by solving a system of linear equations which is known as "ordinary kriging system":

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u}) \gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) - \mu(\mathbf{u}) = \gamma(\mathbf{u}_\alpha - \mathbf{u}) \quad \alpha = 1, \ldots, n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta(\mathbf{u}) = 1 \end{cases} \quad (5)$$

where $\mu(\mathbf{u})$ is the Lagrange parameter accounting for the constraint on the weights. The only information required by the kriging system (5) are semivariogram values for different lags, and these are readily derived once a semivariogram model has been fitted to experimental values. Fig. 5 shows three different types of permissible models that are combined with a nugget-effect model for the fitting of the experimental semivariogram of annual rainfall:

1. Spherical model with range a

$$g(h) = \begin{cases} 1.5\dfrac{h}{a} - 0.5\left(\dfrac{h}{a}\right)^3 & \text{if } h \le a \\ 1 & \text{otherwise} \end{cases} \quad (6)$$

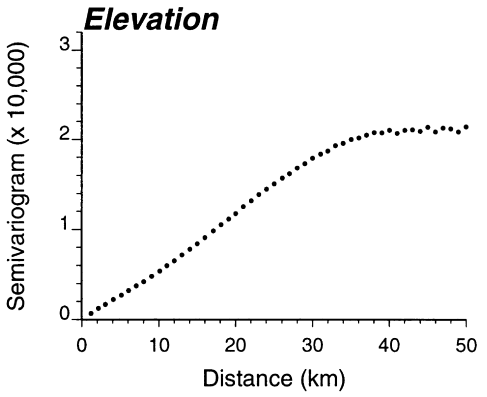

Fig. 6. Experimental semivariogram of elevation computed from the DEM of Fig. 2.

2. Cubic model with range a

$$g(h)=\begin{cases} 7\left(\frac{h}{a}\right)^2-8.75\left(\frac{h}{a}\right)^3+3.5\left(\frac{h}{a}\right)^5 \\ \quad -0.75\left(\frac{h}{a}\right)^7 & \text{if } h \le a \\ 1 & \text{otherwise} \end{cases} \tag{7}$$

3. Dampened hole effect model

$$g(h) = 1.0 - \exp\left(\frac{-3h}{d}\right)\cos\left(\frac{h}{a}\pi\right) \tag{8}$$

where d is the distance at which 95% of the hole effect is dampened out.

The spherical model is the most widely used semivariogram model and is characterized by a linear behavior at the origin. The cubic model (Galli et al., 1984; Chilès and Delfiner, 1999, p. 84) displays a parabolic behavior at the origin and is preferred to the Gaussian model because it avoids numerical unstability in kriging system (Wackernagel, 1998, p. 120). Although no direct information is available on the variability of rainfall over very short distances (the first semivariogram value corresponds to a lag of 5.9 km), ancillary information, such as the more detailed semivariogram of a correlated attribute like altitude (Fig. 6), suggests that one can expect a parabolic behavior for the first lags. Note that a very regular behavior near the origin can be combined with a nugget effect, the latter reflecting measurement errors that are superimposed on the underlying continuous phenomenon. The last type of function is more complex and used to model hole effect (Deutsch and Journel, 1998, p. 26). The three models have been fitted using regression and are such that the weighted sum of squares (WSS) of differences between experimental $\hat{\gamma}(\mathbf{h}_k)$ and model $\gamma(\mathbf{h}_k)$ semivariogram values is minimum:

$$\text{WSS} = \sum_{k=1}^{K} \omega(\mathbf{h}_k)[\hat{\gamma}(\mathbf{h}_k) - \gamma(\mathbf{h}_k)]^2 \tag{9}$$

The weights were taken as $N(\mathbf{h}_k)/[\gamma(\mathbf{h}_k)]^2$ in order to give more importance to the first lags and the ones computed from more data pairs. The cubic model has been retained because it yields the smallest WSS value while being more parsimonious (less parameters to estimate) than the dampened hole effect model. Fig. 4 (bottom graph) shows the rainfall map produced by ordinary kriging using the cubic semivariogram model and the 16 closest observations at each grid node. Like the inverse square distance method, the rainfall map is quite crude, which stresses the importance of accounting for more densely sampled information, such as the digital elevation model of Fig. 2.

3.2. Accounting for elevation

Consider now the situation where the rainfall data $\{z(\mathbf{u}_\alpha), \alpha = 1,\ldots,n\}$ are supplemented by elevation data available at all estimation grid nodes and denoted $y(\mathbf{u})$.

A straightforward approach consists of predicting the rainfall as a function of the co-located elevation, e.g. using a linear relation such as:

$$z^*(\mathbf{u}) = f(y(\mathbf{u})) = a_0^* + a_1^* y(\mathbf{u}) \tag{10}$$

where the two regression coefficients a_0^*, a_1^* are estimated from the set of collocated rainfall and elevation data $\{(z(\mathbf{u}_\alpha), y(\mathbf{u}_\alpha)),\ \alpha = 1,\ldots,n\}$. For example, the relation between annual rainfall and elevation was modeled as $z(\mathbf{u}) = 324.1 + 0.922y(\mathbf{u})$ $R^2 = 0.62$, leading to the rainfall map shown at the top of Fig. 7. A major shortcoming of this type of regression is that the rainfall at a particular grid node $\mathbf{u}$ is derived only from the elevation at $\mathbf{u}$, regardless of the records

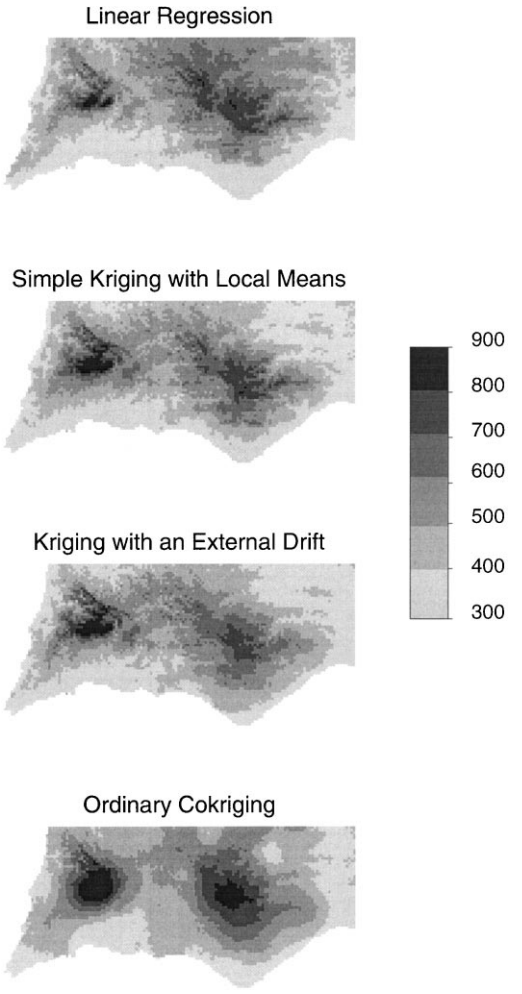

Fig. 7. Annual rainfall maps obtained by interpolation of 36 observations accounting for the digital elevation model of Fig. 2. Four algorithms are considered: linear regression, simple kriging with varying local means, kriging with an external drift and colocated ordinary cokriging.

at the surrounding raingages $\mathbf{u}_\alpha$. Such an approach amounts at assuming that the residual values $r(\mathbf{u}_\alpha) = z(\mathbf{u}_\alpha) - f(y(\mathbf{u}_\alpha))$ are spatially uncorrelated. Spatial correlation of the residuals or of the rainfall observations can be taken into account using the three types of geostatistical algorithms described below.

Simple kriging with varying local means (SKlm) amounts at replacing the known stationary mean in the simple kriging estimate by known varying means $m^*_{\mathrm{SK}}(\mathbf{u})$ derived from the secondary information (Goovaerts, 1997, pp. 190–191):

$$z^*_{\mathrm{SKlm}}(\mathbf{u}) - m^*_{\mathrm{SK}}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{\mathrm{SK}}_\alpha(\mathbf{u})[z(\mathbf{u}_\alpha) - m^*_{\mathrm{SK}}(\mathbf{u}_\alpha)] \tag{11}$$

If the local means are derived using a relation of type (10), the SKlm estimate can be rewritten as the sum of the regression estimate $f(y(\mathbf{u})) = m^*_{\mathrm{SK}}(\mathbf{u})$ and the SK estimate of the residual value at $\mathbf{u}$:

$$z^*_{\mathrm{SKlm}}(\mathbf{u}) = f(y(\mathbf{u})) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{\mathrm{SK}}_\alpha(\mathbf{u}) r(\mathbf{u}_\alpha) \tag{12}$$

where the weights $\lambda^{\mathrm{SK}}_\alpha(\mathbf{u})$ are obtained by solving the simple kriging system:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda^{\mathrm{SK}}_\beta(\mathbf{u}) C_{\mathrm{R}}(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_{\mathrm{R}}(\mathbf{u}_\alpha - \mathbf{u})$$

$$\alpha = 1, \ldots, n(\mathbf{u}) \tag{13}$$

where $C_{\mathrm{R}}(\mathbf{h})$ is the covariance function of the residual RF $R(\mathbf{u}) = Z(\mathbf{u}) - m(\mathbf{u})$, not that of $Z(\mathbf{u})$ itself. If the residuals are uncorrelated, $C_{\mathrm{R}}(\mathbf{h}) = 0 \forall \mathbf{h}$, hence all kriging weights in Eq. (12) are zero and the SKlm estimate is but the value provided by linear regression. Unlike the ordinary kriging system (5), the SK system (13) can be expressed in terms of only covariances because of the lack of constraints on kriging weights (Goovaerts, 1997, p. 135). However, the common practice consists of estimating and modeling the semivariogram $\gamma_{\mathrm{R}}(\mathbf{h})$, then retrieving the covariance $C_{\mathrm{R}}(\mathbf{h})$ as $C_{\mathrm{R}}(0) - \gamma_{\mathrm{R}}(\mathbf{h})$. This conversion from semivariogram to covariance values is automatically performed within most geostatistical softwares, hence the user typically provides only the semivariogram model. For example Fig. 8 shows the semivariogram of residuals for annual rainfall with the model fitted. This model was used to generate the rainfall map in Fig. 7 (2nd graph). The impact of elevation on the rainfall map is less pronounced than for the map generated using linear regression that was but a rescaling of the elevation model (Fig. 7, top graph).

Like the SKlm approach, kriging with an external drift (KED) uses the secondary information to derive the local mean of the primary attribute z, then

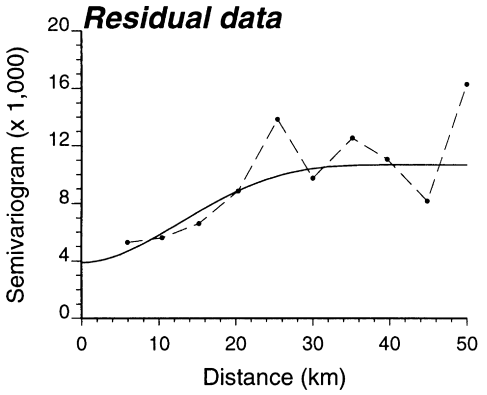

Fig. 8. Experimental residual semivariogram of annual rainfall with the model fitted.

performs simple kriging on the corresponding residuals:

$$z^*_{\mathrm{KED}}(\mathbf{u}) - m^*_{\mathrm{KED}}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{\mathrm{SK}}_{\alpha}(\mathbf{u})[z(\mathbf{u}_\alpha) - m^*_{\mathrm{KED}}(\mathbf{u}_\alpha)] \tag{14}$$

where $m^*_{\mathrm{KED}}(\mathbf{u}) = a^*_0(\mathbf{u}) + a^*_1(\mathbf{u})y(\mathbf{u})$. Estimates (11) and (15) differ by the definition of the local mean or trend. The trend coefficients a^*_0 and a^*_1 are derived once and independently of the kriging system in the SKlm approach, whereas in the KED approach the regression coefficients $a^*_0(\mathbf{u})$ and $a^*_1(\mathbf{u})$ are implicitly estimated through the kriging system within each search neighborhood. In other words, the relation between elevation and rainfall is assessed locally, which allows one to account for changes in correlation across the study area. The coefficients $a^*_0(\mathbf{u})$ and $a^*_1(\mathbf{u})$ can be computed and mapped for interpretation purposes (e.g. see Goovaerts, 1997, p. 201) but they are not required for estimation. Indeed, the usual and equivalent expression for the KED estimate (Wackernagel, 1998, pp. 199–201; Goovaerts, 1997, pp. 194–198) is:

$$z^*_{\mathrm{KED}}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{\mathrm{KED}}_{\alpha}(\mathbf{u})z(\mathbf{u}_\alpha) \tag{15}$$

The kriging weights $\lambda^{\mathrm{KED}}_{\alpha}(\mathbf{u})$ are the solution of the following system of $(n(\mathbf{u}) + 2)$ linear equations:

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda^{\mathrm{KED}}_{\beta}(\mathbf{u})\gamma_{\mathrm{R}}(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu^{\mathrm{KED}}_0(\mathbf{u}) \\ \qquad + \mu^{\mathrm{KED}}_1(\mathbf{u})y(\mathbf{u}_\alpha) = \gamma_{\mathrm{R}}(\mathbf{u}_\alpha - \mathbf{u}) \;\; \alpha = 1,\ldots,n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda^{\mathrm{KED}}_{\beta}(\mathbf{u}) = 1 \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda^{\mathrm{KED}}_{\beta}(\mathbf{u})y(\mathbf{u}_\beta) = y(\mathbf{u}) \end{cases} \tag{16}$$

where $\mu^{\mathrm{KED}}_0(\mathbf{u})$ and $\mu^{\mathrm{KED}}_1(\mathbf{u})$ are 2 Lagrange parameters accounting for the constraints on the weights. Note that elevation data $y(\mathbf{u}_\alpha)$ and $y(\mathbf{u})$ intervene in the kriging system but they are not directly included in the estimate (15). Fig. 7 (third graph) shows the map of KED estimates of annual rainfall, which looks very similar to the SKlm map.

Another approach for incorporating secondary information is cokriging, a multivariate extension of kriging (Goovaerts, 1997, pp. 203–248). When the secondary variable is known everywhere and varies smoothly across the study area (e.g. elevation) there is little loss in retaining in the cokriging system only the secondary datum co-located with the location **u** being estimated (Xu et al., 1992; Goovaerts, 1998). Indeed the co-located elevation $y(\mathbf{u})$ tends to screen the influence of further away elevation data. Moreover, using multiple secondary data can lead to unstable cokriging systems because the correlation between close elevation data is much greater than the correlation between distant rainfall data (Goovaerts, 1997, p. 235). The "co-located" cokriging estimate is then:

$$z^*_{\mathrm{CK}}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda^{\mathrm{CK}}_{\alpha}(\mathbf{u})z(\mathbf{u}_\alpha) + \lambda^{\mathrm{CK}}(\mathbf{u})[\,y(\mathbf{u}) - m_Y + m_Z], \tag{17}$$

where m_Z and m_Y are the global means of the rainfall and elevation data, see Table 1. The second term of Eq. (17) corresponds to a rescaling of the secondary variable (elevation) to the mean of the primary variable (rainfall) to ensure unbiased estimation. The main difference between cokriging and the two previous geostatistical algorithms lies in how the

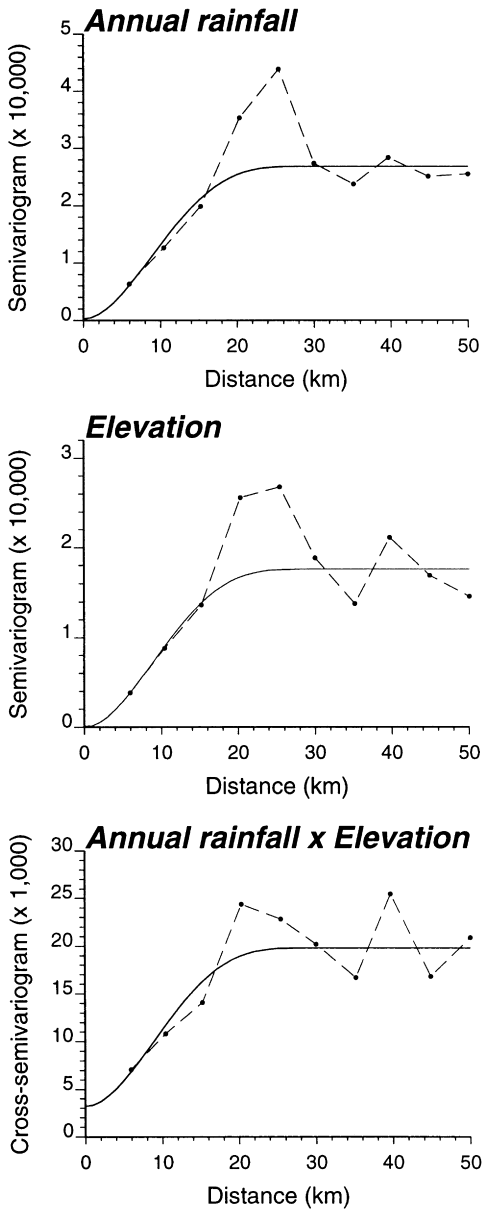

Fig. 9. Experimental semivariograms of annual rainfall and elevation and their cross semivariogram, with the linear model of coregionalization fitted.

elevation value is handled. Whereas that datum directly influences the cokriging estimate, in the SKlm and KED approaches elevation provides information only about the primary trend at location **u**. The cokriging weights are solutions of the following system of $(n(\mathbf{u}) + 2)$ linear equations:

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{\mathrm{CK}}(\mathbf{u})\gamma_{ZZ}(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \lambda^{\mathrm{CK}}(\mathbf{u})\gamma_{ZY}(\mathbf{u}_\alpha - \mathbf{u}) \\ \quad + \mu^{\mathrm{CK}}(\mathbf{u}) = \gamma_{ZZ}(\mathbf{u}_\alpha - \mathbf{u}) \quad \alpha = 1, \ldots, n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{\mathrm{CK}}(\mathbf{u})\gamma_{YZ}(\mathbf{u} - \mathbf{u}_\beta) + \lambda^{\mathrm{CK}}(\mathbf{u})\gamma_{YY}(0) \\ \quad + \mu^{\mathrm{CK}}(\mathbf{u}) = \gamma_{ZY}(0) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{\mathrm{CK}}(\mathbf{u}) + \lambda^{\mathrm{CK}}(\mathbf{u}) = 1 \end{cases}$$

where $\gamma_{ZY}(\mathbf{u}_\alpha - \mathbf{u})$ is the cross-semivariogram value between primary and secondary variables at locations $\mathbf{u}_\alpha$ and **u** respectively. Fig. 9 shows the experimental semivariograms of elevation and annual rainfall, and their cross semivariogram computed as:

$$\hat{\gamma}_{ZY}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})] \times [y(\mathbf{u}_\alpha) - y(\mathbf{u}_\alpha + \mathbf{h})] \quad (18)$$

Note that the semivariogram of elevation has been computed from only the 36 climatic stations, not the entire DEM as in Fig. 6, to avoid possible inconsistencies in the subsequent modeling of direct and cross semivariograms (Goovaerts, 1997, p. 52; Wackernagel, 1998, p. 159) A linear model of coregionalization was fitted using an iterative procedure developed by Goulard (1989). The three semivariograms are modeled as a linear combination of the same set of basic models (here a nugget effect and a cubic model with range 29.18 km) so that the WSS criterion is minimum under the constraints of positive semi-definiteness of the matrix of sills, see Goovaerts (1999) for more details. Fig. 7 (bottom graph) shows the map of cokriging estimates of annual rainfall. Unlike the three previous techniques, the details of the elevation map do not appear in the rainfall map which actually shows more similarities with the ordinary kriging map at the bottom of Fig. 4.

4. Evaluation of the different interpolators

The performances of the seven interpolators were

assessed and compared using cross validation (Isaaks and Srivastava, 1989, pp. 351–368). The idea consists of removing temporarily one rainfall observation at a time from the data set and "re-estimate" this value from remaining data using the alternative algorithms. The comparison criterion is the mean square error (MSE) of prediction which measures the average square difference between the true rainfall $z(\mathbf{u}_\alpha)$ and its estimate $z^*(\mathbf{u}_\alpha)$:

$$\text{MSE} = \frac{1}{n}\sum_{\alpha=1}^{n}[z(\mathbf{u}_\alpha) - z^*(\mathbf{u}_\alpha)]^2 \quad (19)$$

where $n = 36$ for the Algarve data set. The value of this criterion should be close to zero if the algorithm is accurate. For linear regression, the MSE was simply computed as the average square residual value for the linear model fitted using all 36 observations, which means that the prediction error would tend to be underestimated for this method. Although the various kriging interpolators provide an estimate of the error variance, the latter has not been retained as a performance criterion because in practice it usually provides little information on the reliability of the kriging estimate, as reminded by several authors (Journel, 1993; Armstrong, 1994).

The same interpolators as described in previous sections and illustrated for annual rainfall have been applied to monthly rainfall data. Fig. 10 shows, for example, the semivariograms of raw rainfall data and residuals for four of the wettest months (recall Table 1). Because of the control of the relief on the spatial distribution of precipitations, the semivariograms have similar shape although the nugget effect and range of the cubic model fluctuate from one month to another. For the same four months, Fig. 11 shows the maps of ordinary kriging estimates; to facilitate the comparison of grayscale maps, six equally probable classes of values have been created for each month. Despite the similarity of their semivariograms, monthly rainfall maps show slightly different patterns: smaller precipitations are recorded along the West Coast in December and February whereas high precipitation cells move towards the West in April and October.

Fig. 12 shows the mean square errors of prediction produced by each of the seven interpolation algorithms for the monthly (Jan–Dec) and annual (Ann) rainfall. Results are expressed as proportions of the prediction error of the linear regression approach, hence absolute values are easily retrieved by multiplying these percentages by the values of Table 1 (last column). The conclusions are as follows:

1. Larger predictions errors are obtained for the three algorithms that ignore elevation, with the worst results produced by Thiessen polygon. It is noteworthy that for several months, and on average over the year, ordinary kriging yields smaller prediction errors than linear regression of rainfall against elevation.
2. Except for the period from June to September which is characterized by low rainfall amounts (Table 1), multivariate geostatistical algorithms perform better than linear regression which disregards the information provided by surrounding climatic stations.
3. Among geostatistical algorithms that account for elevation, simple kriging with varying local means and kriging with an external drift yield slightly better results than the more complex ordinary cokriging.

To identify the factors that might be responsible for these relative prediction performances, ratios of MSE values have been plotted against parameters, such as the relative nugget effect of the rainfall semivariogram, the relative nugget effect of the residual semivariogram, and the correlation coefficient between rainfall and elevation. The first scattergram (Fig. 13, left top graph) shows that the benefit of using ordinary kriging instead of the inverse square distance method (i.e. a larger MSE ratio Inv/OK) decreases as the spatial dependence between observations weakens, which is indicated by a larger relative nugget effect for the rainfall semivariogram. Similarly, the benefit of using a geostatistical approach (SKlm) instead of linear regression to account for elevation decreases as the spatial dependence between residual data weakens (larger nugget effect for the residual semivariogram), see Fig. 13 (right top graph).

The two middle graphs of Fig. 13 show the impact of the strength of the correlation between elevation and rainfall on the relative performances of ordinary kriging versus cokriging (left graph) and ordinary kriging versus linear regression (right graph). Note that the smallest correlation coefficients observed

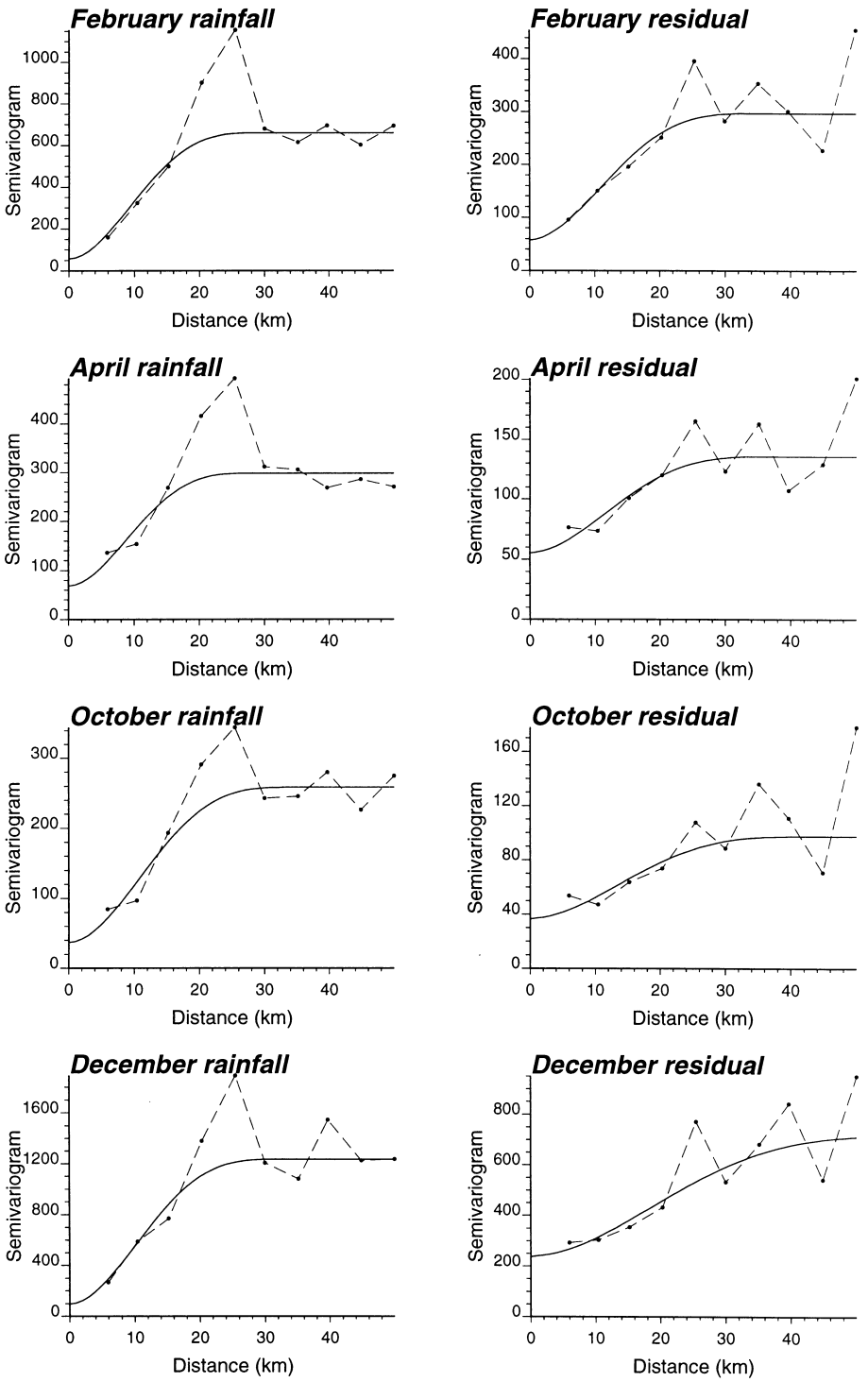

Fig. 10. Experimental semivariograms of monthly rainfall before and after substraction of the local means provided by linear regression of rainfall against elevation at the 36 climatic stations, with the model fitted using weighted least-square regression.

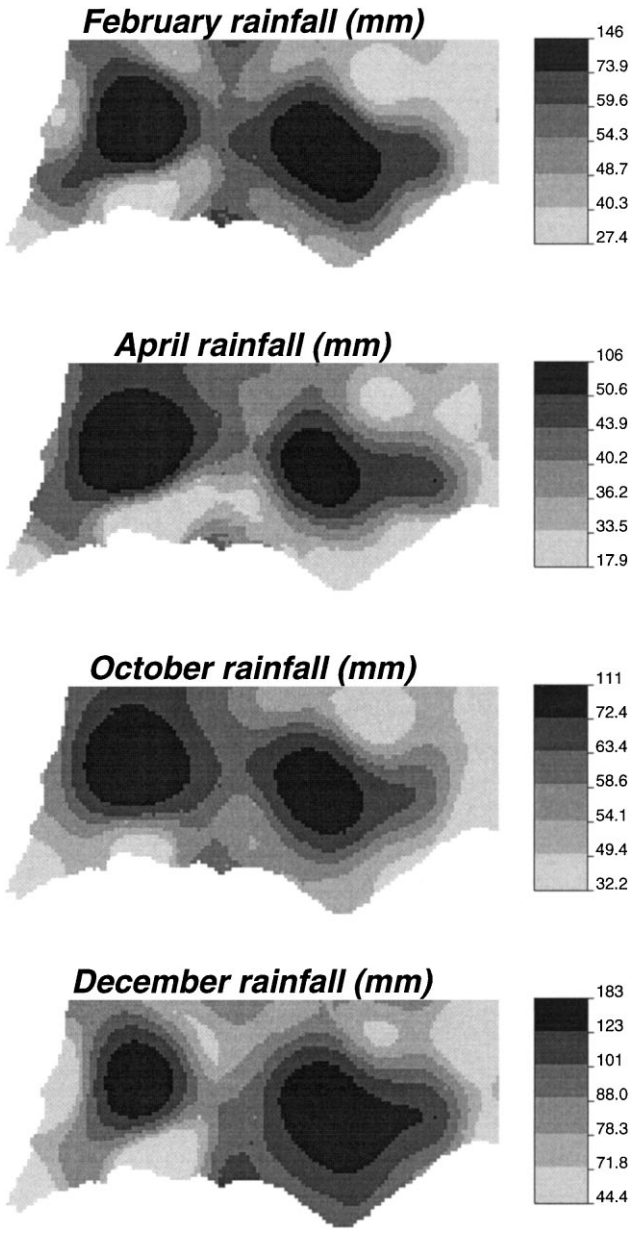

Fig. 11. Monthly rainfall maps obtained by ordinary kriging of 36 observations using the semivariogram models of Fig. 10 (left column).

during the summer have not been taken into account because they also correspond to negligible amounts of rainfall, see Table 1. The gain of ordinary cokriging versus kriging increases as elevation brings more information on rainfall (higher correlation coefficient). The other graph indicates that ordinary kriging performs better than linear regression (ratio smaller than 1) when the correlation between elevation and rainfall is moderate ($\rho < 0.75$) and so the information from surrounding stations is worth integration. For larger correlation coefficients, the elevation at the estimation grid node screens the influence of rainfall data

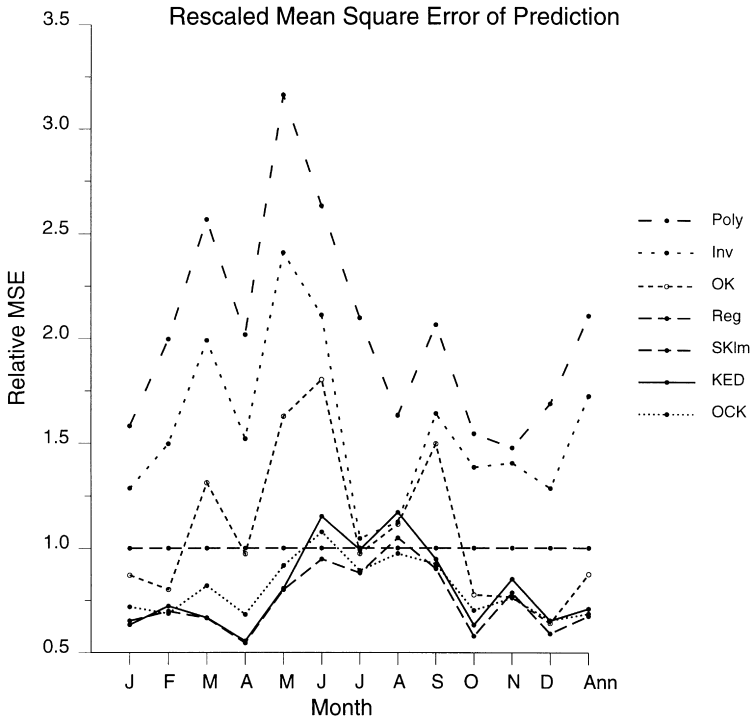

Fig. 12. Mean square error of prediction produced by each of the seven interpolation algorithms for monthly (Jan–Dec) and annual (Ann) rainfall. Results are expressed as proportions of the prediction error of the linear regression approach.

at surrounding sites and so spatial information is less valuable.

Goovaerts (1997, pp. 217–221), showed that the contribution of the secondary information to the cokriging estimate depends not only on the correlation between primary and secondary variables, but also on their patterns of spatial continuity. As the relative nugget effect of the primary semivariogram increases, the increasingly noisy primary data carry less information and the secondary data have larger weight, in particular when the cross semivariogram and the semivariogram of the secondary variable have a small relative nugget effect. As the semivariogram of elevation has typically a small nugget effect (recall Fig. 6), one can expect similar results, as confirmed by the bottom graph of Fig. 13. The gain of ordinary cokriging versus kriging increases as the spatial dependence between observations weakens, which is indicated by a larger relative nugget effect for the rainfall semivariogram (as for the middle graphs July and August results are not included).

5. Conclusions

Our results confirm previous findings (e.g. Creutin and Obled, 1982) that for low-density networks of raingages geostatistical interpolation outperforms techniques, such as the inverse square distance or Thiessen polygon, that ignore the pattern of spatial dependence which is usually observed for rainfall data: the mean square error of kriging prediction is up to half the error produced using inverse square distance. Prediction can be further improved if correlated secondary information, such as a digital elevation model, is taken into account. This paper has reviewed different ways to incorporate such exhaustive secondary information, and cross validation has shown that prediction performances can vary greatly among algorithms.

The most straightforward approach consists of deriving the rainfall value directly from the colocated elevation through a (non)linear regression. By so doing, one ignores however the information provided

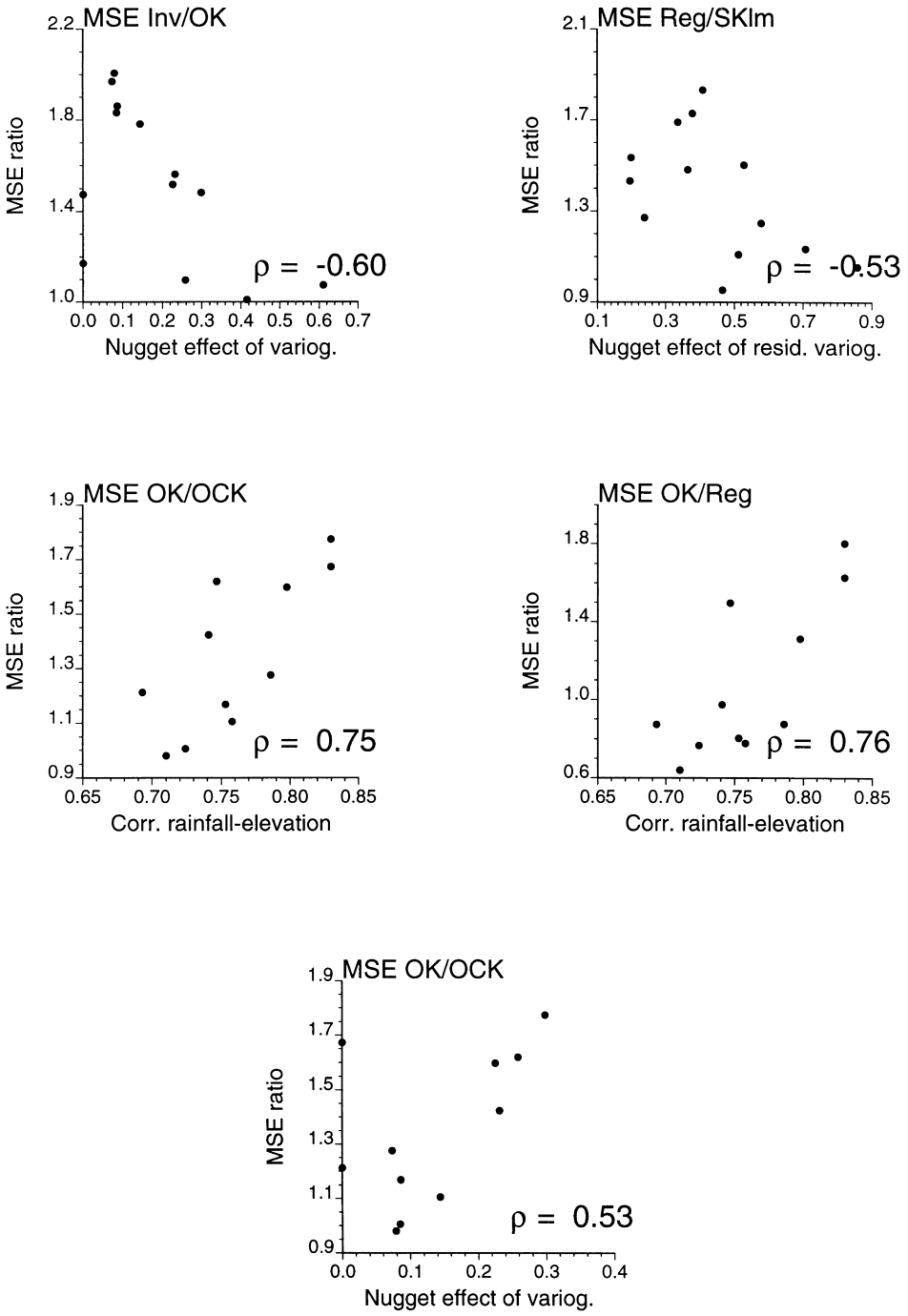

Fig. 13. Scattergrams between the ratio of mean square errors of prediction for different interpolation techniques and parameters, such as the relative nugget effect of the rainfall semivariogram, the relative nugget effect of the residual semivariogram, and the correlation coefficient between rainfall and elevation.

by surrounding climatic stations which is critical when the correlation between the two variables is not too strong and when the residuals are spatially correlated. In this case study, ordinary kriging which ignores elevation is in fact better than linear regression when the correlation is smaller than 0.75! An easy way to account for both elevation and spatial correlation is to interpolate the regression residuals using geostatistics, that is to use simple kriging with varying local means (SKlm). For most months SKlm provides the smallest mean square error of prediction and so performs better that the more sophisticated kriging with an external drift (KED) that evaluates the correlation between elevation and rainfall within each search neighborhood. The last technique is cokriging that interpolates the rainfall as a linear combination of surrounding rainfall observations and the colocated elevation. This approach is the most demanding because three semivariograms must be inferred and jointly modeled, a task that is however alleviated by the recent development of automatic fitting procedures. Cokriged maps show less details than the SKlm and KED maps that are greatly influenced by the pattern of the DEM. In this case study, the additional complexity of cokriging does not pay off in that the prediction errors are not smaller than the ones provided by SKlm and KED.

Further research should investigate whether other environmental descriptors, such as the distance to the sea or the slope orientation, allow one to explain a larger proportion of the spatial variability displayed by rainfall. Whereas cokriging and kriging with multiple external drifts may become very cumbersome to apply, SKlm provides an easy way to incorporate several secondary variables and, for this data set, it yields the best prediction. In this case study, accounting for elevation using multivariate geostatistical algorithms (SKlm, KED and OCK) generally reduces the OK prediction error as long as the correlation coefficient is larger than 0.75. A similar correlation threshold has been reported by Asli and Marcotte (1995) who further concluded that the introduction of secondary information in estimation seems worthy only for correlations above 0.4. The benefit of multivariate techniques can therefore become marginal if the correlation between rainfall and elevation (or other environmental descriptors) is too small, as it might be the case for rainfall accumulations during shorter time steps. Besides the correlation between elevation and rainfall it is also important to look at their patterns of spatial continuity. Elevation data that are moderately correlated with rainfall (i.e. a correlation between 0.4 and 0.7) but exhibit a much smaller relative nugget effect than the rainfall semivariogram may still improve prediction using cokriging, in particular if the nugget effect of the cross semivariogram between rainfall and elevation is small.

Acknowledgements

The author thanks Mr. Nuno de Santos Loureiro for the Algarve data set.

References

Armstrong, M., 1994. Is research in mining geostats as dead as a dodo? In: Dimitrakopoulos, R. (Ed.)., Geostatistics for The Next Century, Kluwer Academic, Dordrecht, pp. 303–312.

Asli, M., Marcotte, D., 1995. Comparison of approaches to spatial estimation in a bivariate context. Math. Geol. 27 (5), 641–658.

Azimi-Zonooz, A., Krajewski, W.F., Bowles, D.S., Seo, D.J., 1989. Spatial rainfall estimation by linear and non-linear cokriging of radar-rainfall and raingage data. Stochastic Hydrol. Hydraul. 3, 51–67.

Bedient, P.B., Huber, W.C., 1992. Hydrology and Floodplain Analysis. 2nd ed, Addison-Wesley, Reading, MA.

Borga, M., Vizzaccaro, A., 1997. On the interpolation of hydrologic variables: formal equivalence of multiquadratic surface fitting and kriging. J. Hydrol. 195 (1–4), 160–171.

Chilès, J.-P., Delfiner, P., 1999. Geostatistics. Modeling Spatial Uncertainty. Wiley, New York.

Creutin, J.D., Obled, C., 1982. Objective analyses and mapping techniques for rainfall fields: an objective comparison. Water Resour. Res. 18 (2), 413–431.

Creutin, J.D., Delrieu, G., Lebel, T., 1988. Rain measurement by raingage-radar combination: a geostatistical approach. J. Atm. Oceanic Tech. 5 (1), 102–115.

Daly, C., Neilson, R.P., Phillips, D.L., 1994. A statistical topographic model for mapping climatological precipitation over montainous terrain. J. Appl. Meteor. 33 (2), 140–158.

Deutsch, C.V., Journel, A.G., 1998. GSLIB: Geostatistical Software Library and User's Guide. 2nd ed, Oxford University Press, New York.

Dirks, K.N., Hay, J.E., Stow, C.D., Harris, D., 1998. High-resolution studies of rainfall on Norfolk Island Part II: interpolation of rainfall data. J. Hydrol. 208 (3-4), 187–193.

Galli, A., Gerdil-Neuillet, F., Dadou, C., 1984. Factorial kriging analysis: a substitute to spectral analysis of magnetic data. In: Verly, G., David, M., Journel, A.G., Maréchal, A. (Eds.), Geostatistics for Natural Resources Characterization, Reidel, Dordrecht, pp. 543–557.

Goovaerts, P., 1997. Geostatistics for Natural Resources Evaluation. Oxford University Press, New York.

Goovaerts, P., 1998. Ordinary cokriging revisited. Math. Geol. 30 (1), 21–42.

Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. Geoderma 89, 1–46.

Goulard, M., 1989. Inference in a coregionalization model. In: Armstrong, M. (Ed.), Geostatistics, Kluwer Academic, Dordrecht, pp. 397–408.

Hevesi, J.A., Flint, A.L., Istok, J.D., 1992a. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part I: structural analysis. J. Appl. Meteor. 31, 661–676.

Hevesi, J.A., Flint, A.L., Istok, J.D., 1992b. Precipitation estimation in mountainous terrain using multivariate geostatistics. Part II: Isohyetal maps. J. Appl. Meteor. 31, 677–688.

Isaaks, E.H., Srivastava, R.M., 1989. An Introduction to Applied Geostatistics. Oxford University Press, New York.

Journel, A.G., 1993. Geostatistics: roadblocks and challenges. In: Soares, A. (Ed.), Geostatistics Tróia '92, Kluwer Academic, Dordrecht, pp. 213–224.

Journel, A.G., Huijbregts, C.J., 1978. Mining Geostatistics. Academic Press, New York.

McCuen, R.H., 1998. Hydrologic Analysis and Design. 2nd ed, Prentice Hall, Englewood Cliffs, NJ.

Phillips, D.L., Dolph, J., Marks, D., 1992. A comparison of geostatistical procedures for spatial analysis of precipitations in mountainous terrain. Agr. Forest Meteor. 58, 119–141.

Raspa, G., Tucci, M., Bruno, R., 1997. Reconstruction of rainfall fields by combining ground raingauges data with radar maps using external drift method. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong '96, Kluwer Academic, Dordrecht, pp. 941–950.

Tabios, G.Q., Salas, J.D., 1985. A comparative analysis of techniques for spatial interpolation of precipitation. Water Resour. Bull. 21 (3), 365–380.

Thiessen, A.H., 1911. Precipitation averages for large areas. Monthly Weather Rev. 39 (7), 1082–1084.

Wackernagel, H., 1998. (completely revised). Multivariate Geostatistics. 2nd ed, Springer, Berlin.

Xu, W., Tran, T.T., Srivastava, R.M., Journel, A.G., 1992. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. Society of Petroleum Engineers, paper no. 24742.